1 **Table of Contents**


15
16

# 6.0 Test Method Reliability

An assessment of test method reliability (intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM 2003). Intralaboratory reproducibility refers to the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol. Interlaboratory reproducibility refers to the extent to which different laboratories can replicate results using the same protocol and test substances, and indicates the extent to which a test method can be transferred successfully among laboratories.

This section describes the reliability assessment for the BG1Luc ER TA test method, which was based on validation study results for substances tested multiple times within and across laboratories.

## 6.1 Intralaboratory Reproducibility

As discussed in **Section 4.2**, the agonist and antagonist DMSO control and antagonist E2 control RLU values were the only quantitative values used for acceptance criteria for agonist test plates throughout the study, therefore intralaboratory reproducibility of the BG1Luc ER TA agonist and antagonist test methods was assessed by comparing: 1) RLU values for the agonist and antagonist DMSO control and the antagonist E2 control for all plates tested within each laboratory during the course of the validation study and 2) results from Phase 2a and 2b testing during which 12 substances were tested in at least three independent experiments in each of the three laboratories.

### 6.1.1 Agonist DMSO Control

Because DMSO control RLU values are not normalized, they vary considerably between test plates and across time. Therefore, intralaboratory reproducibility was evaluated by comparing the within plate variability of the DMSO control RLU values for all test plates that passed acceptance criteria (i.e., CV associated with within plate DMSO control RLU values). The range of means and CV values for within plate DMSO control RLU values are provided in **Table 6-1** (see **Annex L** for the mean and CV of individual agonist test plates). Although mean plate DMSO RLU values ranged from a low of 511 and a high of 9885, with a mean of 3749, within plate variability of DMSO RLU control values between replicate DMSO wells was low with CV values ranging from 1% to 43% with a mean of 8%. Of the 218 agonist test plates that passed acceptance criteria,

only 6 plates had within plate CV values greater that 20% (see **Annex L** for individual test plate mean DMSO control RLU values and associated CV values).

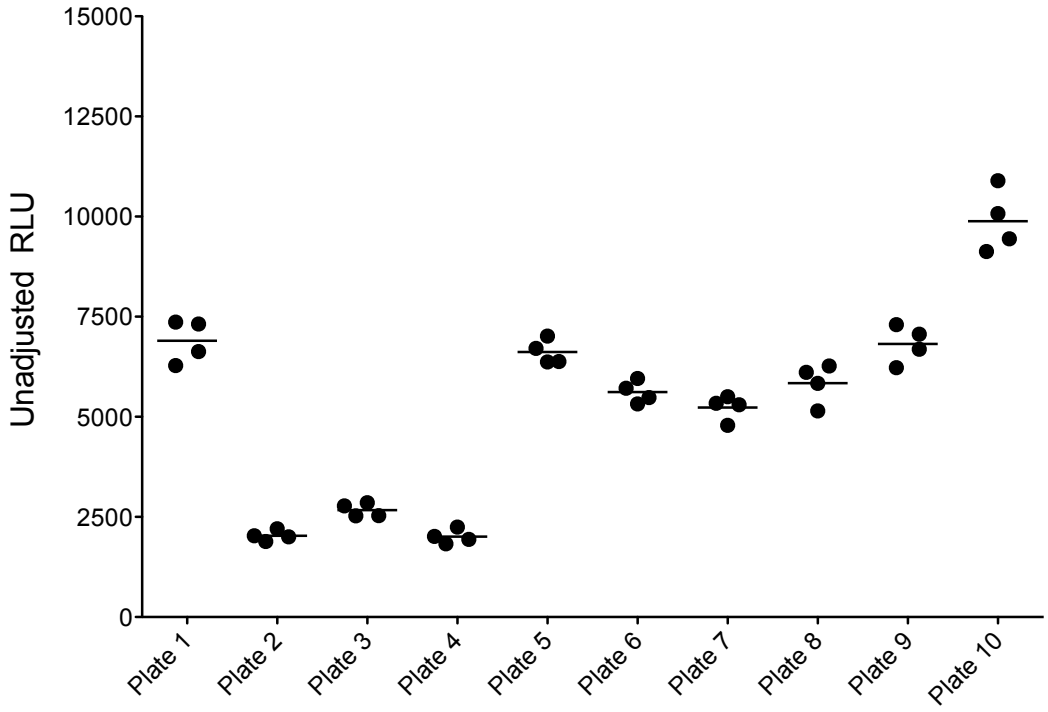
**Table 6-1 Agonist Within Plate DMSO Control Data**

| Laboratory | Mean and Range of DMSO Control RLU Values | Mean and Range of CV (%) | N |
|---|---|---|---|
| XDS | 2800 (511-9885) | 8 (1-43) | 93 |
| ECVAM | 3379 (828-7306) | 8 (1-33) | 60 |
| Hiyoshi | 5465 (1362-9383) | 6 (1-24) | 65 |
| All Laboratories | 3749 (511-9885) | 8 (1-43) | 218 |

Abbreviations: CV = coefficient of variation; ECVAM = European Centre for the Validation of Alternative Methods; Hiyoshi = Hiyoshi Corporation; N = number of plates that passed acceptance criteria; XDS = Xenobiotic Detection Systems, Inc.

**Figures 6-1** through **6-3** provides the within-plate agonist DMSO control RLU values for Phase 1 of the validation study as an example of the low variability for this parameter. As discussed above, within-plate CVs were low throughout the validation study.

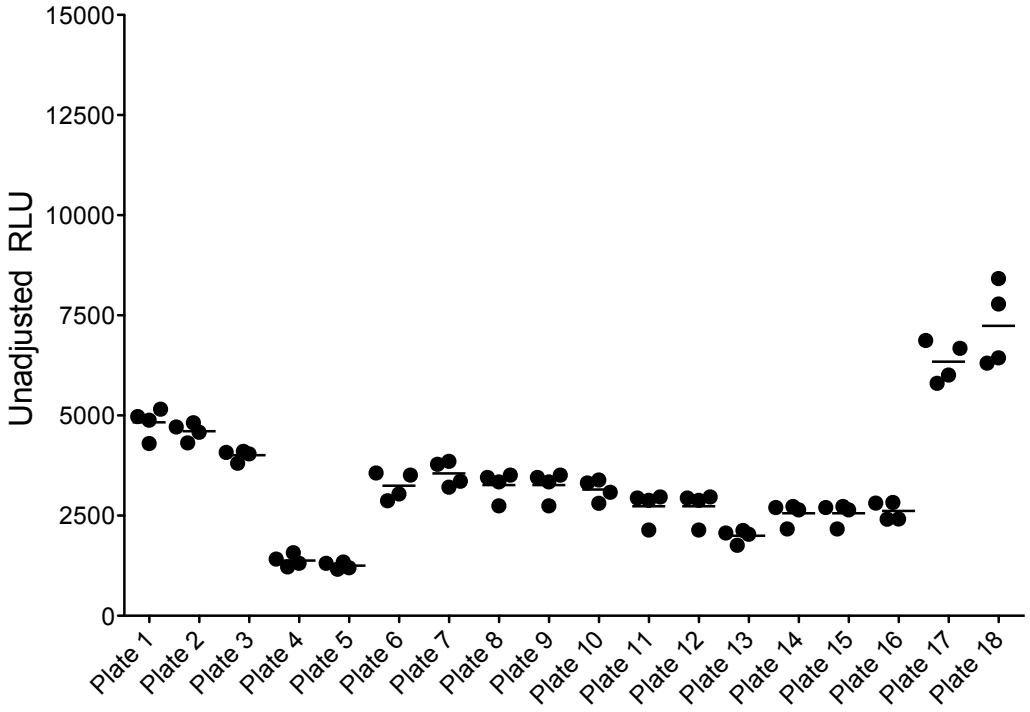
**Figure 6-1 Agonist DMSO Control Within-Plate RLU Values during Phase 1 at XDS[a,b]**



[a]Each point represents the non-normalized DMSO value for a single well of a 96 well plate.

[b]Within-plate DMSO variance at XDS during Phase 1 was fairly low, with coefficients of variation ranging from 5% to 9%.

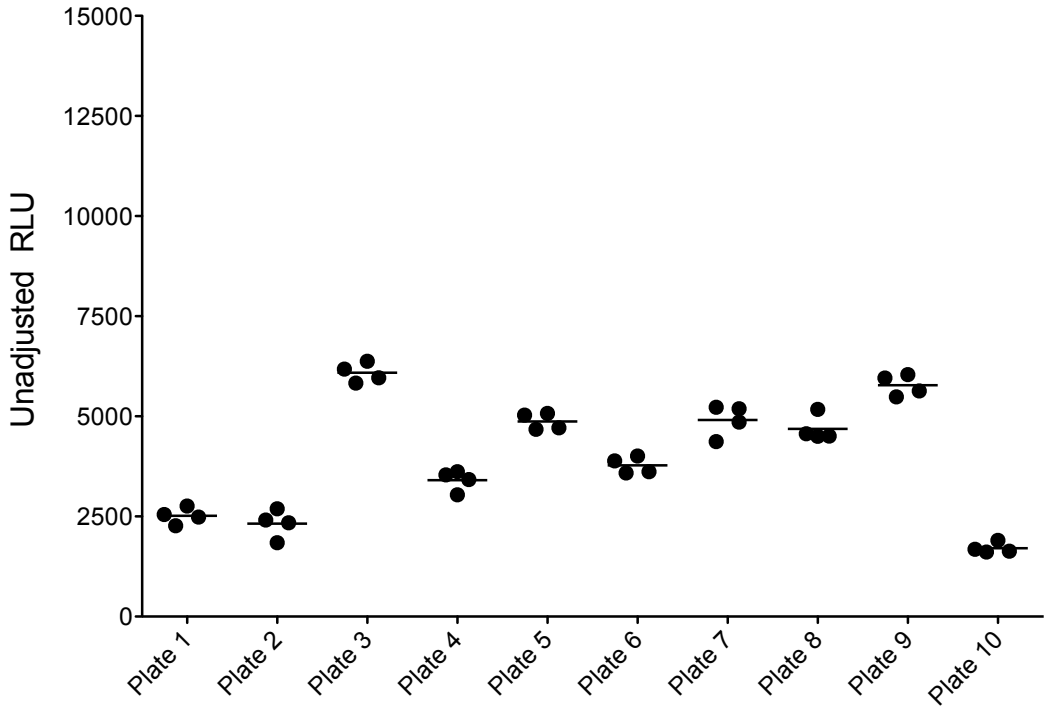
**Figure 6-2 Agonist DMSO Control Within-Plate RLU Values during Phase 1 at ECVAM[a,b]**




[a]Each point represents the non-normalized DMSO value for a single well of a 96 well plate.

[b]Within-plate DMSO variance during Phase 1 was fairly low, with coefficients of variation ranging from 2% to 14%.

**Figure 6-3 Agonist DMSO Control Within-Plate RLU Values during Phase 1 at Hiyoshi[a,b]**




[a]Each point represents the non-normalized DMSO value for a single well of a 96 well plate.

[b]Within-plate DMSO variance during Phase 1 was fairly low, with coefficients of variation ranging from 4% to 15%.

### 6.1.2 Agonist E2 Reference Standard $EC_{50}$ and Methoxychlor Control

Although E2 reference standard $EC_{50}$ and methoxychlor control RLU values were not used for plate acceptance following Phase 2a of the validation study (see **Sections 2.7.1**), these values were collected throughout the study for information purposes and the means and SDs for these parameters from all plates that passed acceptance criteria are provided in **Table 6-2**.

**Table 6-2 Agonist E2 $EC_{50}$ and Methoxychlor Control Values**

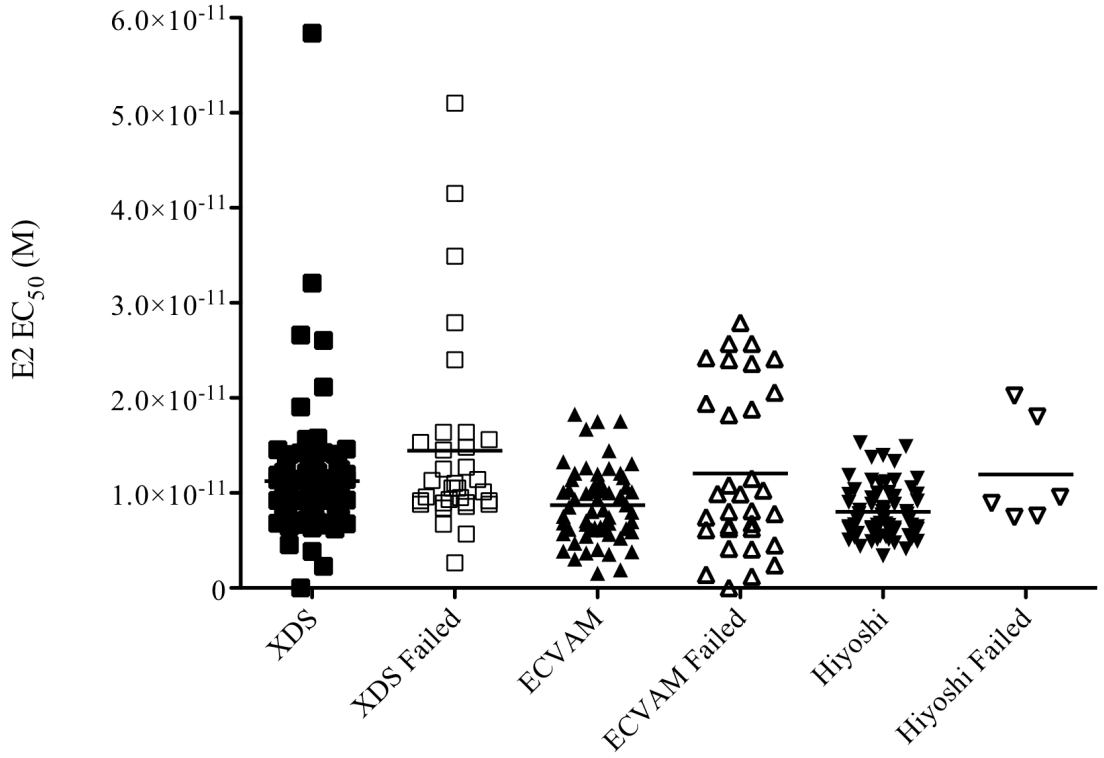
| Laboratory | Mean | SD | N |
|---|---|---|---|
| **E2 Reference Standard $EC_{50}$ (M** | | | |
| XDS | $1.1 \times 10^{-11}$ | $6.7 \times 10^{-12}$ | 93 |
| ECVAM | $1.1 \times 10^{-11}$ | $1.9 \times 10^{-11}$ | 60 |
| Hiyoshi | $8.0 \times 10^{-12}$ | $2.8 \times 10^{-12}$ | 65 |
| **Methoxychlor (RLU)** | | | |
| XDS | 6075 | 1283 | 93 |
| ECVAM | 6246 | 1609 | 60 |
| Hiyoshi | 8029 | 1233 | 65 |

Abbreviations: $EC_{50}$ =the half maximal effective concentration; ECVAM = European Centre for the Validation of Alternative Methods; Hiyoshi = Hiyoshi Corporation; N = number of plates that passed acceptance criteria; SD = standard deviation; XDS = Xenobiotic Detection Systems, Inc.

As indicated in **Table 6-2**, mean E2 reference standard $EC_{50}$ values ranged between $8.0 \times 10^{-12}$ to $1.2 \times 10^{-11}$ M. Methoxychlor control RLU values, which ranged from 6152 to 8029 were highest at Hiyoshi and lowest at XDS.

E2 reference standard $EC_{50}$ and methoxychlor control RLU values for all plates tested during the validation study are presented in **Figures 6-4** through **6-5**. Laboratories are relatively consistent when data from only acceptable plates are considered. These data also indicated that the variability of each parameter is generally higher when considering only values obtained from plates that failed one or more acceptance criteria. With the exception of E2 $EC_{50}$ at XDS, all outlier values among the parameters evaluated were associated with these failed plates.

**Figure 6-4 Agonist E2 Reference Standard $EC_{50}$ Values[a,b,c,d]**



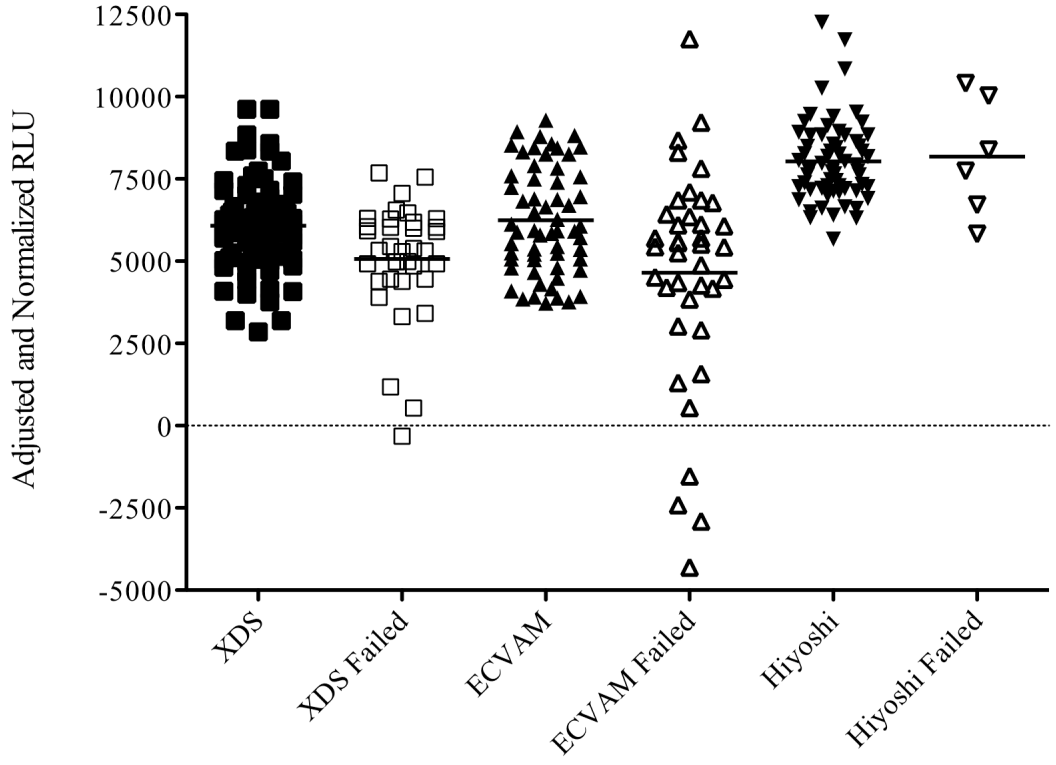

[a]Each point represents a single plate.

[b]An $EC_{50}$ value ($1.18 \times 10^{-9}$M) from one experiment that failed acceptance criteria at XDS was excluded from the graph.

[c]$EC_{50}$ values ($1.69 \times 10^{-10}$M, and $7.78 \times 10^{-11}$M) from two experiments that passed acceptance criteria at XDS were excluded from the graph to minimize scale distortion.

[d]An $EC_{50}$ value ($1.56 \times 10^{-10}$M) from one experiment that passed acceptance criteria at ECVAM was excluded from the graph to minimize scale distortion.

**Figure 6-5 Agonist Methoxychlor Control Values[a,b,c]**




[a]Each point represents a single plate.
[b]Methoxychlor control values (35581, -74511, and –6995) from three experiments that failed acceptance criteria at XDS were excluded from this graph to minimize scale distortion.
[c]Methoxychlor control values (-127587, and –8464) from two experiments failed acceptance criteria at ECVAM were excluded from the graph to minimize scale distortion.

### 6.1.3 Intralaboratory Reproducibility of Phase 2 Agonist Reference Substances

As described in **Section 2.0**, test substances are assigned as positive or negative for agonist activity based on a specific set of criteria. The resulting classifications for each of the 12 substances that were tested at least three times at each laboratory were used to evaluate the extent of intralaboratory agreement (see **Table 6-3**). Although the classifications for some of the test substances differed among the laboratories, there was 100% agreement within each laboratory for each of the three repeat tests. There were no "inadequate" data generated at any lab during this phase of the validation study.

**Table 6-3 Intralaboratory Agreement for Multiple Testing of 12 Phase 2 Agonist Substances Tested Independently Three Times at Each Laboratory**

| Activity per Test | XDS | ECVAM | Hiyoshi |
|---|---|---|---|
| Agreement Within Laboratory | 12/12 (100%) | 12/12 (100%) | 12/12 (100%) |
| +++[a] | 8/12 | 12/12 | 9/12 |
| −−−[b] | 4/12 | 0/12 | 3/12 |
| Discordance Within Laboratory | 0/12 (0%) | 0/12 (0%) | 0/12 (0%) |
| ++−[c] | 0/12 | 0/12 | 0/12 |
| +−−[d] | 0/12 | 0/12 | 0/12 |

Abbreviations: + = positive test result; - = negative test result

[a]+++ indicates that each of three replicate tests within each laboratory had a classification as positive.

[b]--- indicates that each of three replicate tests within each laboratory had a classification as negative

[c]++- indicates that in two of three replicate tests, a test substance was classified as positive. The substance was classified as negative in a third replicate test

[d]+--- indicates that in one of three replicate tests, the test substance was classified as positive. The substance was classified as negative in the remaining two tests.

### 6.1.4 Antagonist DMSO Control

Because DMSO control RLU values are not normalized, they vary considerably between test plates and across time. Therefore, intralaboratory reproducibility was evaluated by comparing the within plate variability of the DMSO control RLU values for all test plates that passed acceptance criteria (i.e., CV associated with within plate DMSO control RLU values). The range of means and CV values for within plate DMSO control RLU values are provided in **Table 6-4** (see **Annex L** for the mean and CV of individual antagonist test plates). Although mean plate DMSO RLU values ranged from a low of 132 and a high of 8451, with a mean of 3299, within plate variability of DMSO RLU control values between replicate DMSO wells was low with CV values ranging from 1% to 52% with a mean of 8%. Of the 194 agonist test plates that passed acceptance criteria, only 8 plates had within plate CV values greater that 20% (see **Annex L** for individual test plate mean DMSO control RLU values and associated CV values).

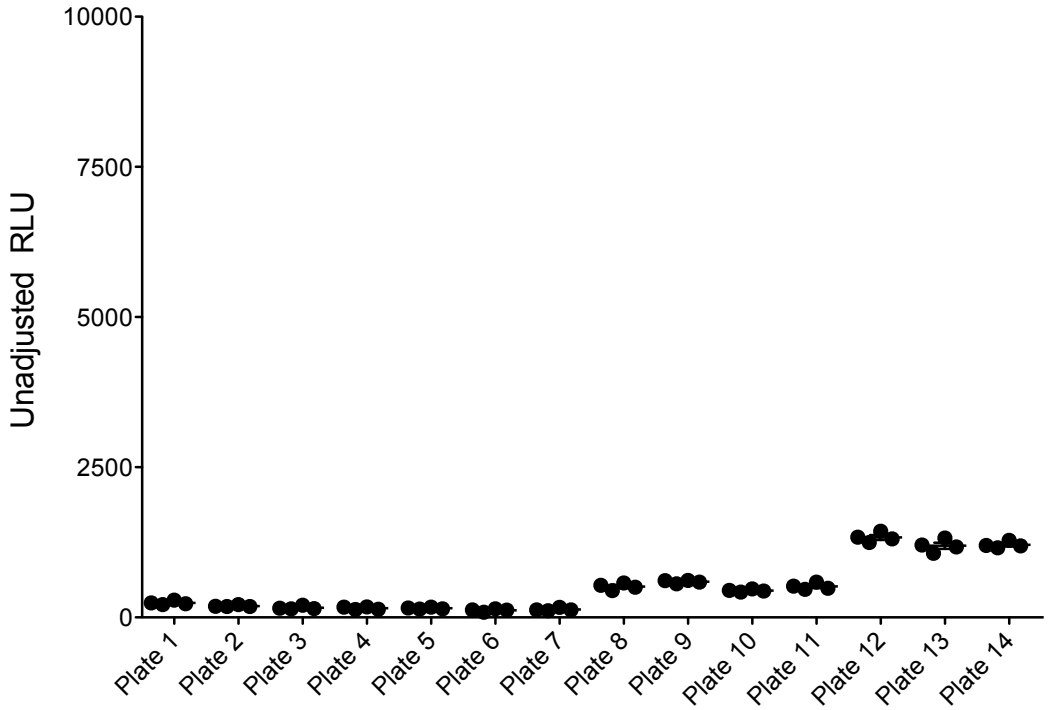
**Table 6-4 Antagonist DMSO Control Values**

| Laboratory | Mean and Range of DMSO Control RLU Values | Mean and Range of CV (%) | N |
|---|---|---|---|
| XDS | 2230 (132-6860) | 9 (1-52) | 79 |
| ECVAM | 3622 (1352-7333) | 9 (1-37) | 62 |
| Hiyoshi | 4030 (1625-8451) | 6 (1-20) | 53 |
| All Laboratories | 3299 (132-8451) | 8 (1-52) | 194 |

Abbreviations: CV = coefficient of variation; ECVAM = European Centre for the Validation of Alternative Methods; Hiyoshi = Hiyoshi Corporation; N = number of plates that passed acceptance criteria; XDS = Xenobiotic Detection Systems, Inc.

**Figures 6-6** through **6-8** provides the within-plate agonist DMSO control RLU values for Phase 1 of the validation study as an example of the low variability for this parameter. As discussed above, within plate CVs were low throughout the validation study.

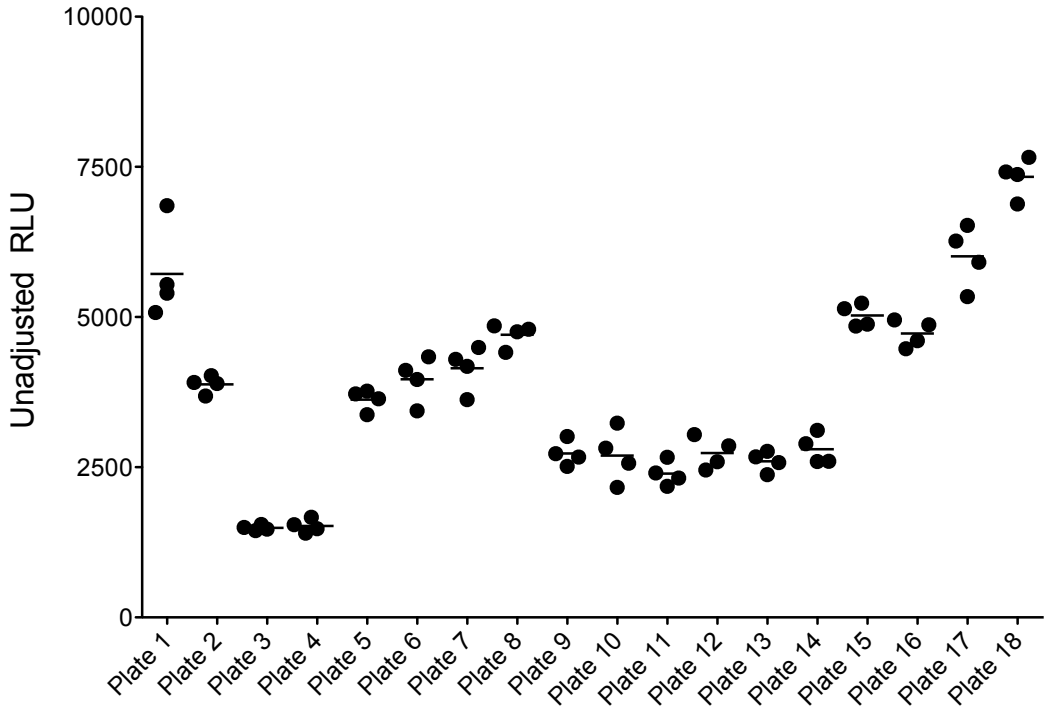
**Figure 6-6 Antagonist DMSO Control Within-Plate RLU Values during Phase 1 at XDS[a,b]**




[a]Each point represents the non-normalized DMSO value for a single well of a 96 well plate.

[b]Within-plate DMSO variance during Phase 1 was fairly low, with coefficients of variation ranging from 3% to 18%.

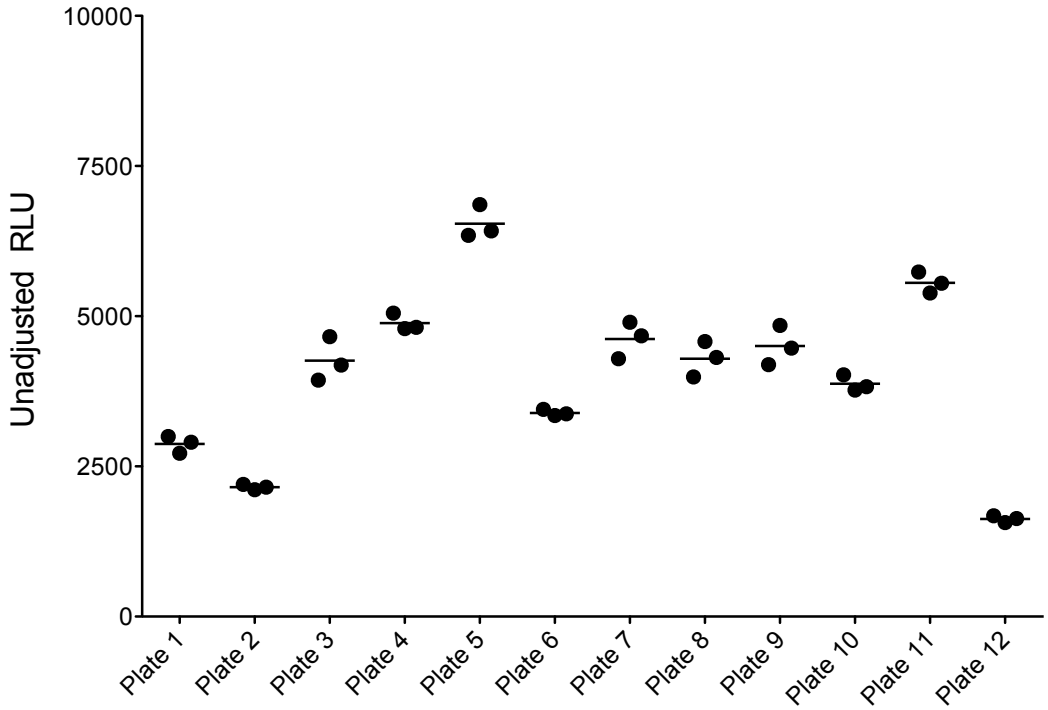
**Figure 6-7 Antagonist DMSO Control Within-Plate RLU Values during Phase 1 at ECVAM[a,b]**




[a]Each point represents the non-normalized DMSO value for a single well of a 96 well plate.

[b]Within-plate DMSO variance during Phase 1 was fairly low, with coefficients of variation ranging from 3% to 17%.

**Figure 6-8 Antagonist DMSO Control Within-Plate RLU Values during Phase 1 at Hiyoshi[a,b]**




[a]Each point represents the non-normalized DMSO value for a single well of a 96 well plate.

[b]Within-plate DMSO variance during Phase 1 was fairly low, with coefficients of variation ranging from 3% to 9%.

### 6.1.5 Antagonist E2 Control

Normalized and adjusted antagonist E2 control RLU values were used as an acceptance criterion throughout the validation study. The mean, SD, and CV calculated for the E2 control RLU value from all antagonist test plates that passed acceptance criteria are provided in **Table 6-5**). Mean E2 control RLU values ranged from 5793 at Hiyoshi to 9246 at ECVAM and variability was low with associated CV values ranging from 9% at ECVAM and 19% at XDS.

**Table 6-5 Antagonist E2 Control Values**

| Laboratory | Mean | SD | CV (%) | N |
|---|---|---|---|---|
| XDS | 7524 | 1443 | 19 | 79 |
| ECVAM | 9246 | 805 | 9 | 62 |
| Hiyoshi | 5793 | 791 | 14 | 53 |

Abbreviations: CV = coefficient of variation; ECVAM = European Centre for the Validation of Alternative Methods; Hiyoshi = Hiyoshi Corporation; N = number of plates that passed acceptance criteria; SD = standard deviation; XDS = Xenobiotic Detection Systems, Inc.

### 6.1.6 Antagonist Ral Reference Standard $IC_{50}$ and Flavone Control

Although Ral reference standard $IC_{50}$ and flavone control RLU values were not used for plate acceptance following Phase 2a of the validation study (see **Sections 2.7.2**), these values were collected throughout the study for information purposes and the means and SDs for these parameters from all plates that passed acceptance criteria are provided in **Table 6-6**.

**Table 6-6 Antagonist Ral $IC_{50}$ and Flavone Control Values**

| Laboratory | Mean | SD | N |
|---|---|---|---|
| **Ral Reference Standard $IC_{50}$(M)** | | | |
| XDS | $1.1 \times 10^{-9}$ | $5.6 \times 10^{-10}$ | 79 |
| ECVAM | $1.3 \times 10^{-9}$ | $5.6 \times 10^{-10}$ | 62 |
| Hiyoshi | $1.2 \times 10^{-9}$ | $2.9 \times 10^{-10}$ | 53 |
| **Flavone (RLU)** | | | |
| XDS | 3774 | 1366 | 79 |
| ECVAM | 599 | 468 | 62 |
| Hiyoshi | 873 | 772 | 53 |

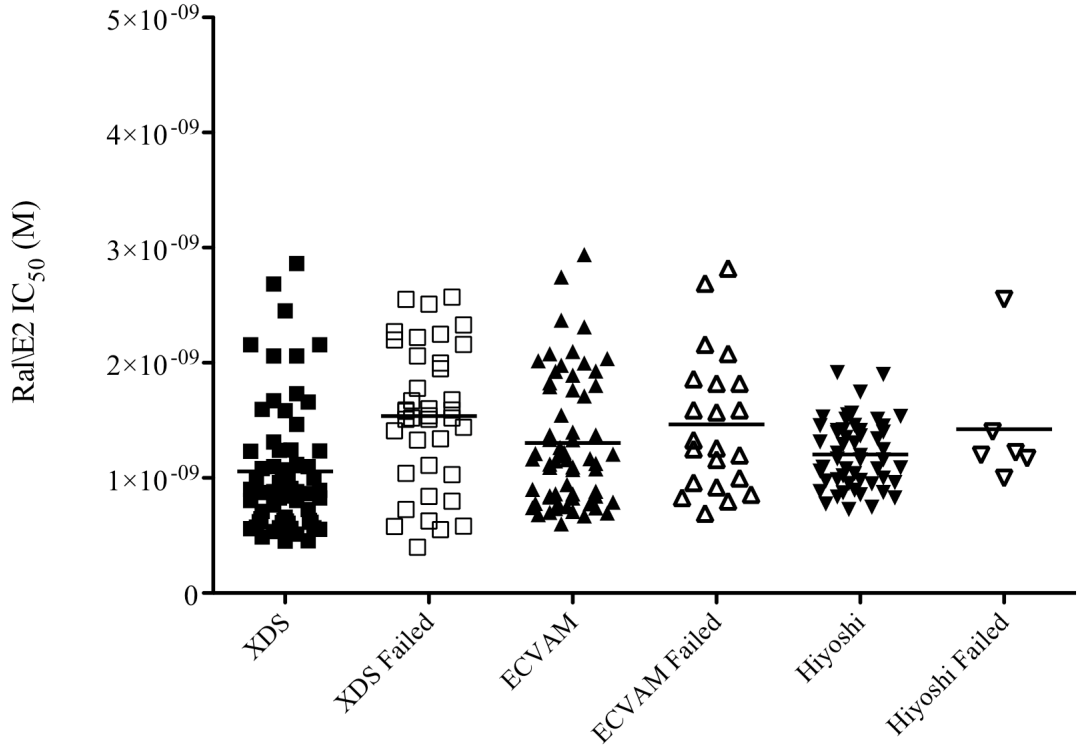
Abbreviations: $IC_{50}$ - half maximal inhibitory concentration; ECVAM = European Centre for the Validation of Alternative Methods; Hiyoshi = Hiyoshi Corporation; N = number of plates that passed acceptance criteria; SD = standard deviation; XDS = Xenobiotic Detection Systems, Inc.

As indicated in **Table 6-3**, Ral reference standard $IC_{50}$ values ranged between $1.1 \times 10^{-9}$ to $1.3 \times 10^{-9}$ M and flavone control RLU values ranged between 599 at Hiyoshi to 3774 at XDS.

Ral reference standard $IC_{50}$, flavone control, and E2 control RLU values for all plates tested during the validation study are presented in **Figures 6-9** through **6-11**. Laboratories are relatively consistent when data from only acceptable plates are considered. These data also indicated that the variability of each parameter is generally higher when considering only values obtained from

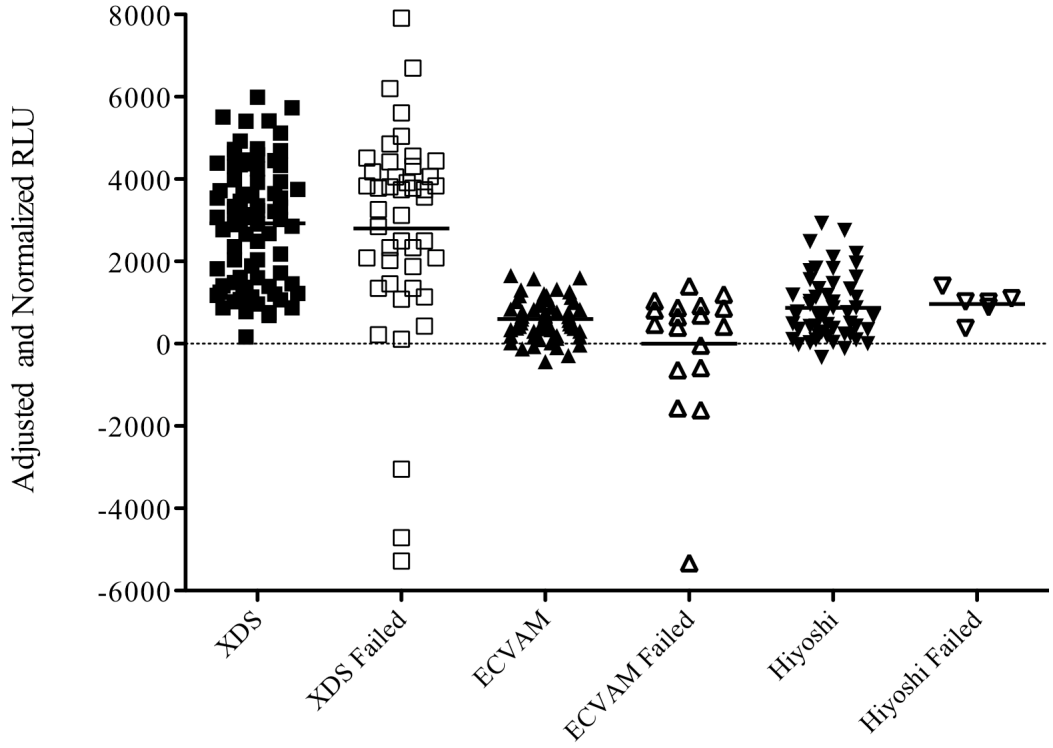
plates that failed one or more acceptance criteria. Additionally, any outlier values among the parameters evaluated were associated with these failed plates.

**Figure 6-9 Antagonist Ral Reference Standard $IC_{50}$ Values[a]**




[a]Each point represents a single plate.

**Figure 6-10 Antagonist Flavone Control Values[a,b,c]**

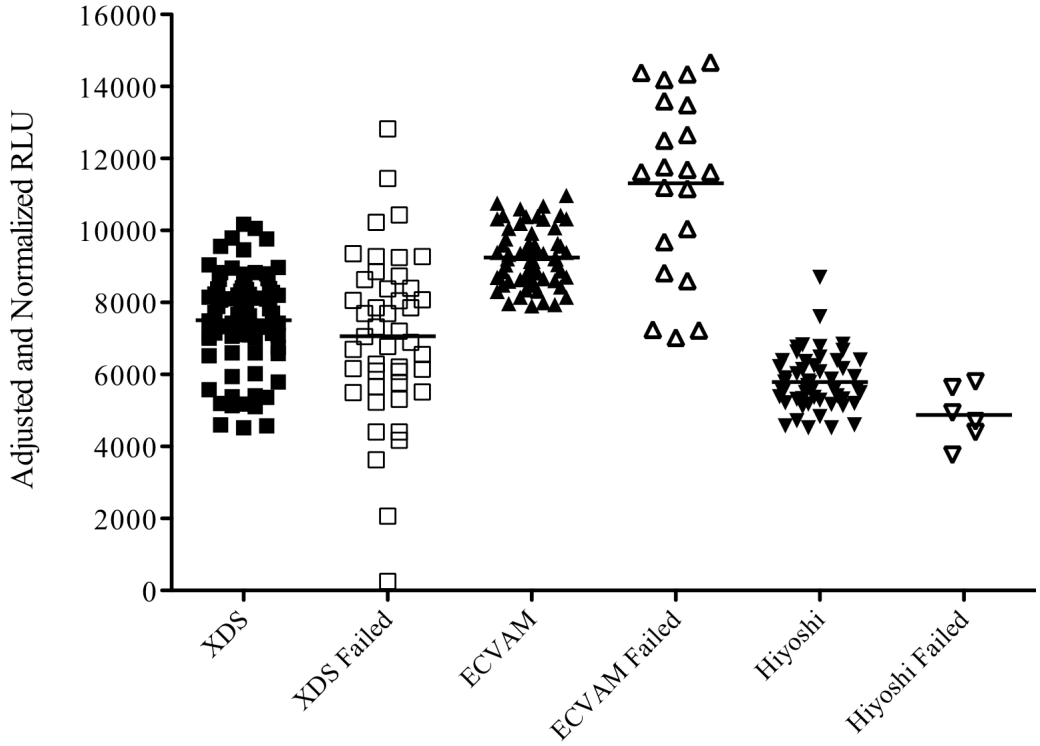



[a]Each point represents a single plate.

[b]Flavone control values from two experiments that passed acceptance criteria at XDS were excluded from the graph (237690, and 23164).

[c]Flavone control values from four experiments that failed acceptance criteria at XDS were excluded from the graph (22676, -21568, –16714, and -8081).

**Figure 6-11 Antagonist E2 Control Values[a,b,c]**




[a]Each point represents a single plate.

[b]E2 control values from two experiments that failed acceptance criteria at XDS were excluded from the graph (41227, and –3995).

[c]A flavone control value from one experiment that failed acceptance criteria at ECVAM was excluded from the graph (20345).

### 6.1.7 Intralaboratory Reproducibility of Phase 2 Antagonist Reference Substances

As described in **Section 2.0**, test substances are assigned as positive or negative for antagonist activity based on a specific set of criteria. The resulting classifications for each of the 12 substances that were tested at least three times at each laboratory were used to evaluate the extent of intralaboratory agreement (see **Table 6-7**). Although the classifications for some of the test substances differed among the laboratories, there was 100% agreement within each laboratory for each of the three repeat tests. There were no "inadequate" data generated at any lab during this phase of the validation study.

**Table 6-7 Intralaboratory Agreement for Multiple Testing of 12 Phase 2 Antagonist Substances Tested Independently Three Times at Each Laboratory**

| Activity per Test | XDS | ECVAM | Hiyoshi |
|---|---|---|---|
| Agreement Within Laboratory | 12/12 (100%) | 12/12 (100%) | 12/12 (100%) |
| +++[a] | 2/12 | 2/12 | 2/12 |
| −−−[b] | 10/12 | 10/12 | 10/12 |
| Discordance Within Laboratory | 0/12 (0%) | 0/12 (0%) | 0/12 (0%) |
| ++−[c] | 0/12 | 0/12 | 0/12 |
| +−−[d] | 0/12 | 0/12 | 0/12 |

Abbreviations: + = positive test result; - = negative test result

[a]+++ indicates that each of three replicate tests within each laboratory had a classification as positive.

[b]--- indicates that each of three replicate tests within each laboratory had a classification as negative

[c]++- indicates that in two of three replicate tests, a test substance was classified as positive. The substance was classified as negative in a third replicate test

[d]+--- indicates that in one of three replicate tests, the test substance was classified as positive. The substance was classified as negative in the remaining two tests.

## 6.2 Interlaboratory Reproducibility

Similar to the intralaboratory analyses described in **Sections 6.1.3** and **6.1.7**, the classifications for each of the substances that were tested for agonist and antagonist activity during Phases 2 and 3 were also used to evaluate the extent of interlaboratory agreement as an indicator of reproducibility among the laboratories.

### 6.2.1 Interlaboratory Reproducibility of Phase 2 Reference Substances

For each of the 12 substances that were tested at least three times for agonist and antagonist activity during Phase 2, agreement among the three laboratories was determined based on the consensus classification assigned by each laboratory for each of the 12 substances (see **Table 4-11** and **4-12** for agonist and antagonist results respectively). As previously noted, there were no "inadequate" data generated at any laboratory during this phase of the validation study.

As indicated in **Table 6-8**, all three laboratories classified the same eight of twelve (67%) substances as agonists (positive). Among the remaining four substances, one (flavone) was identified as positive by 2/3 laboratories (ECVAM and Hiyoshi), but negative at XDS. Although the starting concentrations for flavone were identical at all three laboratories (100 μg/mL), all three tests at XDS were uniformly negative and there was no increasing concentration response noted. The other three substances that were discordant among the laboratories (corticosterone, vinclozolin, and atrazine) were identified as negative by 2/3 laboratories (XDS and Hiyoshi), but positive at ECVAM. It is noted that the all three substances appeared to be negative for agonist activity when range finder tested at ECVAM but all three substances were uniformly positive when comprehensively tested. Therefore, the positive agonist results observed for corticosterone, vinclozolin, and atrazine during comprehensive testing at ECVAM may be due to contamination of stocks subsequent to range finder testing.

**Table 6-8 Interlaboratory Agreement for Phase 2 Test Substances**

| Results Among Laboratories | Agonist Testing | Antagonist Testing |
|---|---|---|
| Agreement Among Laboratories | 8/12 (67%) | 12/12 (100%) |
| +++ | 8/12 | 2/12 |
| −−− | 0/12 | 10/12 |
| Discordance Among Laboratories | 4/12 (33%) | 0/12 (0%) |
| ++− | 1/12 | 0/12 |
| +−− | 3/12 | 0/12 |

Abbreviations: + = positive test result; - = negative test result

[a]+++ indicates that the substance was classified as positive at all three laboratories.

[b]--- indicates that substance was classified as negative at all three laboratories

[c]++- indicates that in two of three laboratories a test substance was classified as positive. The substance was classified as negative in the third laboratory

[d]+--- indicates that in one of three laboratories, the test substance was classified as positive. The substance was classified as negative in the third laboratory

Among the substances tested for antagonist activity, there was 100% agreement among the three laboratories for all 12 substances. Two of these substances (dibenzo[*a*,*h*]anthracene and tamoxifen) were positive in all three laboratories. The other 10 substances were negative in all three laboratories (see **Table 6-8)**.

### 6.2.2 Interlaboratory Reproducibility of Phase 3 Agonist Reference Substances

The classifications for each of the 41 substances that were tested once for agonist activity at all three laboratories during Phase 3 were also used to evaluate the extent of interlaboratory agreement. Unlike Phase 2, some of the substances tested in Phase 3 produced results that were considered inadequate (i.e., substances that failed to meet the decision criteria for either a positive or negative response defined in **Section 2.7.1.4**). While such results could not be used in the evaluation of test method accuracy detailed in **Section 5.0**, these results are tabulated in this section as an indication of how often one or more laboratories produced inadequate results. However, for the purposes of an interlaboratory reproducibility assessment, only those substances that produced a definitive result in at least two of the three laboratories were used.

Of the 41 substances tested in Phase 3, 88% (36/41) produced a definitive result in at least two laboratories, and were therefore used for the assessment of reproducibility. A definitive result (i.e., determination of a positive or negative response) was not determined for the remaining 12% of substances (testing produced results that were inadequate for these substances in at least two laboratories, so were not used for the assessment of interlaboratory reproducibility as noted above). Among these 36 substances, the three laboratories agreed on 83% (30/36) of the substances tested for agonist activity (see **Table 6-9)**. Of the 30 substances that had 100% agreement across laboratories, 20 were positive for ER agonist activity and 10 were negative for ER agonist activity. There was discordance among the laboratories for the remaining six substances, as indicated in the lower portion of **Table 6-9**. Three of these substances (dicofol, fluoranthene, and 2-*sec*-butylphenol) were positive in 2/3 laboratories (XDS and Hiyoshi), but negative in 1/3 laboratory (ECVAM). The other three substances (4-androstenedione, clomiphene citrate, and resveratrol) were discordant between the two laboratories that produced a definitive result (i.e., a negative result produced in one laboratory, a positive result produced in another laboratory, and an inadequate result was produced in the third laboratory).

The discordance among the laboratories for at least four of the six substances listed above (fluoranthene, 2-sec-butylphenol, androstenedione, and resveratrol) appears to have resulted from differences in the concentration selected for comprehensive testing by the discordant laboratory. As detailed in **Section 2.0**, the starting concentration for comprehensive testing is chosen based

on data from range finding tests; the highest dose used for range finding tests is directed related to the highest soluble concentration. For one of these four substances (fluoranthene), the discordance among laboratories appears to be due to differing interpretations of test substance solubility, where the highest concentration used for comprehensive testing at ECVAM is at least an order of magnitude lower than the highest concentration selected at XDS or Hiyoshi (see **Figure 6-12**). For the remaining three substances (2-*sec*-butylphenol, androstenedione, and resveratrol) the differences in starting concentrations for comprehensive testing appear to have resulted from incorrect interpretation of data obtained during range finding experiments (**see Figure 6-13** as an example). The discordance among the laboratories for the remaining two substances (clomifene citrate and dicofol) was not based on either differences in solubility or interpretation of range finder results. Clomifene citrate was clearly positive at Hiyoshi and clearly negative at ECVAM when comprehensively tested over the same concentration range. Although dicofol was positive when tested at Hiyoshi using a starting concentration an order of magnitude higher than XDS and ECVAM, it was clearly positive at XDS and clearly negative at ECVAM when comprehensively tested over the same concentration range.

**Table 6-9 Interlaboratory Agreement for Phase 3 Substances Tested Once at Each Laboratory**

| Results Among Laboratories[a] | Agonist Testing | Antagonist Testing |
|---|---|---|
| Agreement Among Laboratories | 30/36 (83%) | 38/41 (93%) |
| +++[b] | 18/36 | 2/41 |
| −−−[c,d] | 4/36 | 33/41 |
| ++i[e] | 2/36 | 1/41 |
| −−i[f] | 6/36 | 2/41 |
| Discordance Among Laboratories | 6/36 (17%) | 3/41 (7%) |
| ++−[g] | 3/36 | 0/41 |
| +−−[h] | 0/36 | 1/41 |
| +−i[i] | 3/36 | 2/41 |

Abbreviations: + = positive test result; - = negative test result; i = inadequate data

[a]Only those substances that produced a definitive result in at least two of the three laboratories were used in this evaluation. There were five substances that produced an inadequate result in two laboratories during agonist testing and are therefore not included in this table.

[b]+++ indicates that the substance was classified as positive at all three laboratories.

[c]Includes one substance (phenobarbital) that was tested in only two laboratories (XDS and ECVAM, see **Section 3.0**).

[d]--- indicates that substance was classified as negative at all three laboratories

[e]++i indicates that the substance was classified as positive at two of three laboratories, but was inadequate in the third.

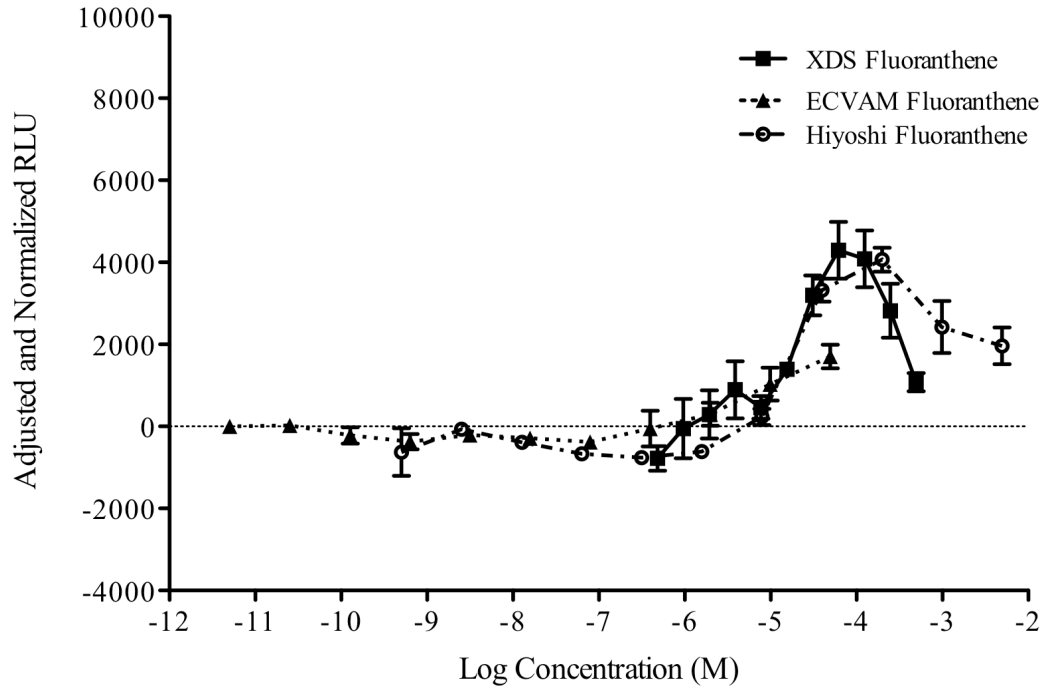
[f]—i indicates that the substance was classified as negative at two of three laboratories, but was inadequate in the third.

[g]++- indicates that in two of three laboratories a test substance was classified as positive. The substance was classified as negative in the third laboratory

[h]+--- indicates that in one of three laboratories, the test substance was classified as positive. The substance was classified as negative in the third laboratory

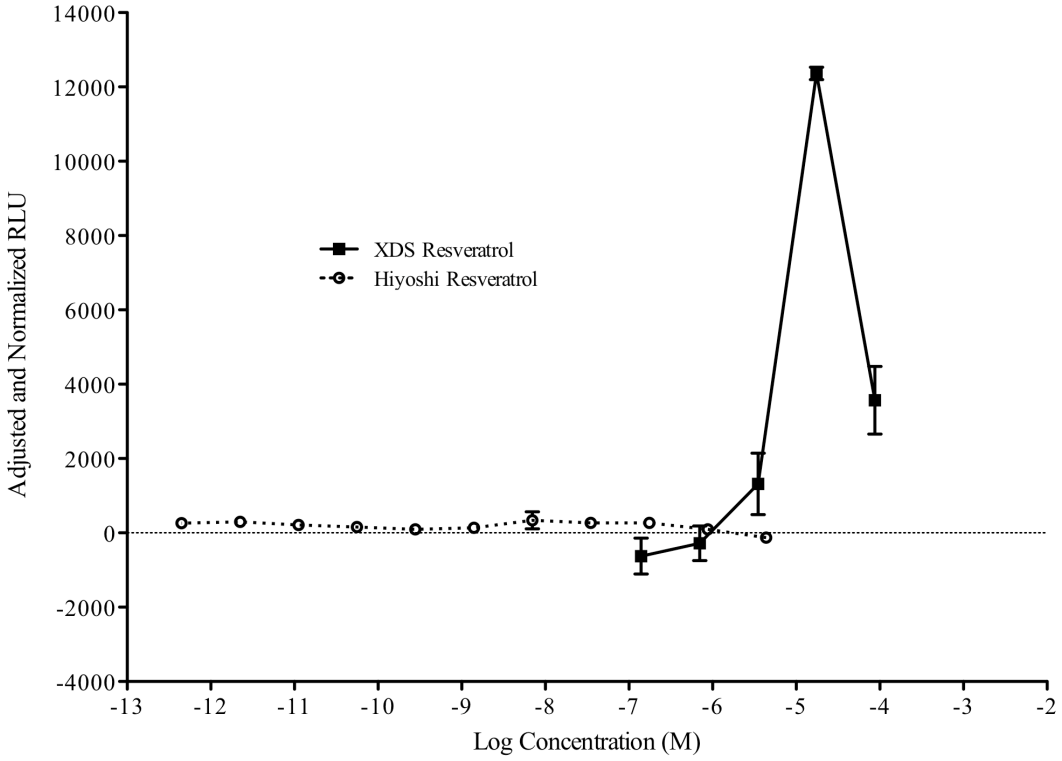
[i]+-i indicates that the subtance was classified as positive at one laboratory, negative at one laboratory, and inadequate at the third laboratory

**Figure 6-12 Fluoranthene Results at All Three Laboratories: Impact of Differences in Solubility on Comprehensive Test Results[a]**




[a]Each point represents the mean adjusted and normalized RLU value and SD from triplicate wells.

**Figure 6-13 Resveratrol Results at XDS and Hiyoshi: Impact of Selecting the Incorrect Starting Concentration Based on Range Finding Results[a]**




[a]Each point represents the mean adjusted and normalized RLU value and SD from triplicate wells; Results for resveratrol at ECVAM were considered inadequate and are therefore not included here.

### 6.2.3 Interlaboratory Reproducibility of Phase 3 Antagonist Reference Substances

The classifications for each of the 41 substances that were tested once for antagonist activity at all three laboratories during Phase 3 were also used to evaluate the extent of interlaboratory agreement. Similar to the Phase 3 agonist test results, some of the substances tested in Phase 3 for antagonist activity produced results that were considered inadequate (i.e., substances that failed to meet the decision criteria for either a positive or negative response defined in **Section 2.7.2.4**). However, unlike the agonist test results, there were no substances tested for antagonist activity that produced inadequate results in more than one laboratory. Therefore, all 41 Phase 3 substances tested for antagonist activity were included in the reproducibility assessment.

The three laboratories agreed on 93% (38/41) of the substances tested for antagonist activity. Most of these substances (85% [35/41]) were identified as negative for antagonist activity; there

were three substances that were positive for antagonist activity. There was discordance among the laboratories for the remaining three substances. One of these substances (diethylstilbestrol) was negative in 2/3 laboratories (XDS and ECVAM), but positive in one laboratory (Hiyoshi). The other two substances (clomiphine citrate and 17α-estradiol) were discordant between the two laboratories that produced a definitive result (i.e., a negative result produced in one laboratory, a positive result produced in another laboratory, and an inadequate result was produced in the third laboratory). It does not appear that any of these three discordant substances can be explained by differences in solubility or interpretation of the range finding data.

If only those substances that produced a definitive result in all three laboratories are considered (n=36), there was 100% agreement for 97% (35/36) of the substances tested. As mentioned previously, substances with "inadequate" data would be retested under the revised testing protocol, and conclusive results would therefore be expected for all test substances. Consequently, the high degree of intralaboratory reproducibility seen when all laboratories produce conclusive results is indicative of the level of performance expected using the revised protocol.

ICCVAM. 2003. ICCVAM Evaluation of In Vitro Test Methods For Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays.National Institute of Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/docs/endo_docs/edfinalrpt0503/edfinrpt.pdf